TRANSACTIONS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 00, Number 0, Pages 000–000
S 0002-9947(XX)0000-0

DIFFERENTIAL EQUIVALENCE CLASSES FOR METRIC PROJECTIONS AND OPTIMAL BACKWARD ERRORS

JOSEPH F. GRCAR

This paper is dedicated to John von Neumann
at the centennial of his birth, on 28th December 1903.

Abstract. It is shown that the sensitivities of metric projections to changes in their sets can be determined by inspecting classes of suitably equivalent functions. The proof depends on parameterizations of theorems related to the mean value and implicit function theorems. The result justifies a common approach to the types of perturbation analyses used in numerical analysis and optimization theory. The method is illustrated by establishing a dependency that John von Neumann suggested among certain errors of numerical calculations, and by establishing the directional differentiability in some cases of the distance between a fixed point and a deforming set.

1. Introduction

Perturbation analysis is the study of small changes among interdependent quantities. If the functional dependencies are not obviously differentiable, then the analyses depend on reasoning specific to the problems on hand. Methods are highly developed for many subjects such as control theory, dynamical systems, numerical analysis, optimization theory, mathematical physics, and various technical fields.

This paper advances a common approach to the kinds of perturbation analyses that occur in numerical analysis and optimization theory. The idea is to group functions with similar perturbational properties into equivalence classes. To answer a perturbational question about a function, one examines its equivalence class to find another function for which the question is more easily answered. This has ever been the practice for differentiable functions, whose perturbations are investigated by replacing them by their linear tangents.

The focus of this paper is the perturbation analysis of a specific kind of function called a metric projection. These are ubiquitous in mathematics so it is natural that they should underlie many perturbational questions. An interesting twist is that the perturbations of interest are those of the set that defines the metric projection rather than of the point that is being projected.

Received by the editors . . . Prepared April 2, 2003.

2000 *Mathematics Subject Classification.* Primary 26B05; Secondary 26B10, 41A29, 41A50, 65G99, 90C31.

Key words and phrases. Collocated level sets, differential equivalence, directional differentiability, implicit function theorem, matrix lower bound, mean value theorem, metric projection, optimal backward errors, parametric optimization, perturbation analysis, sensitivity analysis.

Specifically, this paper identifies some functions that are equivalent in a perturbational sense to the distance between a point and a deformable set. The set is implicitly defined by equality constraints in a finite dimensional Banach space, and it is deformable by a parameterization of its implicit definition; the point is a member of the undeformed set. These limitations are consistent with the more studied case of a fixed set and a variable point, which has been examined in this journal and the *Proceedings* over several years. The proofs depend on parameterized versions of theorems related to the mean value theorem and to the implicit function theorem; these supporting results may have independent interest.

The final part of the paper applies the method of perturbation analysis to answer two questions in numerical analysis and mathematical programming. It establishes the directional differentiability of the distance function, and estimates optimal backward errors. These applications motivate the work. The question in numerical analysis is related to some lesser-known work of John von Neumann.

This is the plan of the paper. The remainder of this introduction describes the two applications. Section 2 introduces the equivalence relations for perturbations. Sections 3, 4, and 5 identify members of the equivalence classes for metric projections with deformable sets. Section 6 returns to the applications in optimization theory and numerical analysis. Section 7 lists open questions.

1.1. **An Optimization Problem.** A point y_1 that attains the distance from a given point y_0 to a closed set S,

$$\operatorname{dist}(y_0, S) = \min_{y \,\in\, S} \|y - y_0\| \,,$$

is called a metric projection of y_0 onto S, and is written $y_1 = P_S(y_0)$. Some authors use this notation for the set of nearest points rather than for a selected nearest point. Metric projections characterize convex sets [37] [38], and they are the most general best approximation problems, so there is broad interest in their differential properties [23].

Differentiability with respect to y_0 for a fixed set S is the more studied case. The large literature can be regarded as exploring situations that are complementary to a basic negative result: although in Hilbert spaces $P_S(y_0)$ is uniquely defined for sets that are closed and convex [65], Kruskal [26] and Shapiro [48] show it need not be directionally differentiable even in the Euclidean plane.[1] The situations under which some differentiability occurs can be classified several ways, according to whether they involve: (1) points internal or external to S, (2) convex or arbitrary sets, (3) Hilbert or Banach spaces, (4) the metric projection or the distance function, and (5) spaces with finite or unspecified dimension. Many of the 2^5 alternatives have been considered: Table 1 surveys the literature, which Table 2 summarizes for $\operatorname{dist}(y, S)$.

It is generally conceded that it is more difficult to investigate the sensitivity of an optimal value whose feasible set is subject to perturbation [6, p. 278]. Thus, in contrast to the many results for perturbations of y_0, the differentiability of metric projections with respect to S (which is the problem of interest in this paper) is less well understood. The problem is encountered primarily in the sensitivity analysis of mathematical programs, which is surveyed by Bonnans and Shapiro [5] [6] and Levitin [27]. The uses of sensitivity analysis are to interpret mathematical programming models [11], to find optimality conditions [43], and to establish

[1] Shapiro [47] discusses the several definitions that are used for directional differentiability.

Table 1. Differentiability results for $P_S(y_0)$ and $\operatorname{dist}(y_0, S)$, with respect to y_0, for a fixed, closed set S.

(1) External points, $y_0 \notin S$:
  (a) Convex S:
    (i) In Hilbert spaces, Haraux [18] finds a class of sets for which $P_S(y_0)$ is directionally differentiable. Fitzpatrick and Phelps [12] characterize the sets whose metric projections are k times continuously Fréchet differentiable.
    (ii) In Banach spaces, remarkably, Holmes [24, p. 99] shows that $\operatorname{dist}(y_0, S)$ always is continuously Fréchet differentiable in the spaces with differentiable norms. A proof just for Hilbert spaces is due to Moreau [24, p. 88–89] [33, p. 286].
  (b) Arbitrary S:
    (i) In Hilbert spaces, Shapiro [49] finds nearly convex sets for which $P_S(y_0)$ is uniquely defined and directionally differentiable. Clarke, Stern and Wolenski [9] identify the sets S that have envelopes of uniform thickness on which $\operatorname{dist}(\cdot, S)$ is continuously Fréchet differentiable. Poliquin, Rockafellar and Thibault [40] characterize the y_0 and S for which $\operatorname{dist}(\cdot, S)$ is continuously differentiable in a neighborhood of y_0.

(2) Internal points, $y_0 \in S$ (the interesting case is $y_0 \in \operatorname{bd}(S)$):
  (a) Convex S:
    (i) In Hilbert spaces, $P_S(y_0)$ is directionally differentiable always. The earliest proof appears to be Zarantonello's [65, p. 300], see also [31, p. 94] and later [18]. Mignot [32] extends this result to projections defined by nonsymmetric bilinear forms.
    (ii) In Banach spaces, Phelps [39, p. 974] finds sets in certain spaces for which $P_S(y_0)$ is well-defined and directionally differentiable.
  (b) Arbitrary S:
    (i) In finite dimensional Banach spaces, Shapiro [45] characterizes the S for which $\operatorname{dist}(y_0, S)$ is directionally differentiable.

Table 2. Differentiability of $\operatorname{dist}(y_0, S)$ with respect to y_0 for a fixed, closed set S. See Table 1 for references and elaboration.

	$y_0 \notin S$	$y_0 \in S$
convex S	$\operatorname{dist}(y_0, S)$ is continuously Fréchet differentiable (C^1) in Banach spaces with dif. norms	$\operatorname{dist}(y_0, S)$ is directionally differentiable (d.d.) in Hilbert spaces
arbitrary S	points and sets for which $\operatorname{dist}(y_0, S)$ is in C^1 have been characterized in Hilbert spaces	sets for which $\operatorname{dist}(y_0, S)$ is d.d. have been characterized in finite dim. Banach spaces

the convergence of algorithms [31] [39]. Without loss of generality from a practical perspective, the theory typically assumes that the objective functions are at least differentiable. As a result much of the sensitivity analysis of optimization problems is inapplicable to metric projections because the function $\|y - y_0\|$ is not differentiable.

It is possible and instructive to apply the general theory to minimizing squared distances in Hilbert spaces. Bonnans and Shapiro [6, p. 434] consider the following optimization problem (in the notation of this paper),

$$\nu(x) = \min_{y \,:\, F(y,\,x) \,\in\, C} \|y - y_0\|_2^2 \,,$$

under quite general conditions:

(i) $F : Y \times X \to Z$ is twice continuously differentiable,
(ii) X and Z are Banach spaces and Y is a Hilbert space,
(iii) C is a closed convex subset of Z,
(iv) $F(y_0, x_0) = 0$ for some x_0, and
(v) Robinson's constraint qualification [6, p. 65] [42, p. 501],

$$0 \in \operatorname{int}\left(F(y_0, x_0) + D_1 F(y_0, x_0) Y - C\right),$$

where set-arithmetic is pointwise, and $D_1 F$ is the partial Fréchet derivative with respect to the first argument of F.

The theory draws many conclusions from these hypotheses, among them that there is a neighborhood N of x_0 and a continuous function $y : N \to Y$, with $y(x_0) = y_0$, so that $y(x)$ is the unique solution of the minimization problem for $\nu(x)$. Moreover, $y(x)$ is directionally differentiable at x_0, and in particular, the derivative in the direction Δx is the solution of a related optimization problem,

$$\lim_{t \,\to\, 0^+} \frac{\nu(x_0 + t\Delta x) - \nu(x_0)}{t} \;=\; \min_{\Delta y \,:\; DF(y_0,\, x_0)(\Delta y,\, \Delta x) \,\in\, T_C(F(y_0,\, x_0))} \|\Delta y\|_2^2 \,,$$

where $T_C(z)$ is the contingent (Bouligand) cone that is tangent to C at z,

$$T_C(z) = \left\{ \Delta z \in Z : \liminf_{t \,\to\, 0^+} \frac{\operatorname{dist}(z + t\Delta z, C)}{t} = 0 \right\} .$$

Note that if Δy is the directional derivative of $y(x)$, then

$$\nu(x_0 + t\Delta x) = \|y(x_0 + t\Delta x) - y_0\|_2^2 \approx \|y(x_0) + t\Delta y - y_0\|_2^2 = \|t\Delta y\|_2^2$$

so $\|\Delta y\|_2$ is the directional derivative of the un-squared distance, $\nu(x)^{1/2}$.

This result, for metric projections in Hilbert spaces, specializes to finite dimensional vector spaces and equality constraints as follows. Let $F : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^p$ and $C = \{0\}$, so that the minimization problem becomes

$$\mu(x) = \min_{y \,:\, F(y,\,x) \,=\, 0} \|y - y_0\|_2 \,,$$

where $F(y_0, x_0) = 0$ for some y_0. In these circumstances Robinson's condition, (v), is simply the hypothesis that that $D_1 F(y_0, x_0) \in \hom(\mathbb{R}^m, \mathbb{R}^p)$ is onto (equivalently, it has full row rank when realized as a Jacobian matrix). In this notation the conclusion is that the metric projection $\mu(x)$ is defined on a neighborhood of x_0 at which it has the directional derivative,

$$\text{(1)} \qquad \lim_{t \,\to\, 0^+} \frac{\mu(x_0 + t\Delta x) - \mu(x_0)}{t} \;=\; \min_{\Delta y \,:\, DF(y_0,\, x_0)(\Delta y,\, \Delta x) \,=\, 0} \|\Delta y\|_2 \,.$$

The methods of this paper establish equation (1) in the more general situation that the norm is any whatsoever and under the weaker hypothesis that F is only continuously differentiable. Aside from the intrinsic interest of this for metric projections, the conclusion is interesting because it suggests that second-order hypotheses, Bonnans and Shapiro's condition (i), may be unnecessarily strong for the first-order sensitivity analysis of mathematical programming problems.

1.2. **A Problem from Numerical Analysis.**

In 1946, Alan Turing visited John von Neumann at Princeton to exchange ideas for building computing machines upon the end of World War II. An issue they discussed [14, p. 291] proved to be an enduring puzzle: how to determine the net effect of rounding errors on a numerical calculation. Von Neumann's and Turing's thoughts can be found in separate papers, [34] and [61], and entwined in subsequent work.

Von Neumann and his collaborator Herman Goldstine apparently were the first to make an unintuitive observation that became paradigmatic. The data from which a calculation begins must be prepared, by measuring them for example, so "it seems reasonable to take the errors of [preparation] into consideration when we analyze what concept and what degree of approximation [in the final result] are significant" [34, p. 1092]. To that end von Neumann and Goldstine noted that an inaccurately calculated result may be correct for some perturbation of the calculation's initial data. They suggested that the size of such compensating perturbations to the data might be used to assess the calculation's accuracy. Data perturbations that account for calculation errors now are called backward errors [64, p. 3].

The determination of minimal backward errors can be viewed as a metric projection in the following way. In the terminology of numerical analysis, the "data" are y, the "solution" is x, and the numerical problem is defined by $F(y, x) = 0$ where F is the "residual" function. An instance of the numerical problem is given by some data, y_0, for which there is a "true" solution, x_0. Given an approximate solution, $x \approx x_0$, the backward error problem is to find the minimal size $\mu(x)$ of data perturbations $y - y_0$ for which $F(y, x) = 0$,

$$\mu(x) = \min_{y \,:\, F(y,x) \,=\, 0} \|y - y_0\| \, .$$

Formulas for $\mu(x)$ have been derived in many specific cases. Each has an underlying numerical problem, which defines F, and a norm or norms with respect to which the size of minimal perturbations have been found. Table 3 surveys these results. Sun [54, p. 358] observed that the residual seems always to occur in expressions for $\mu(x)$. This apparent dependence of the minimal backward error on the residual was anticipated by von Neumann and Goldstine. While studying matrix inversion, they remarked that bounds for the backward error can be derived from bounds for the residual. "We leave the working out of the details, which can be prosecuted in several different ways, to the reader" [34, p. 1093].

This paper justifies the observations of von Neumann, Goldstine, and Sun in general circumstances. Rather than considering a specific calculation, as is done in numerical analysis, this paper addresses all calculations. A general estimate for the size of minimal backward errors is found that is both optimal in a differential sense and easier to evaluate than the original metric projection. Further, the size of minimal backward errors is shown to have an unique estimate as a norm of the underlying numerical problem's residual.

Table 3. Results for optimal backward errors of specific numerical problems.

Linear equations: Optimal data perturbations were first considered by Oettli and Prager [35] and then by Rigal and Gaches [41] for systems of algebraic linear equations (LE). For example, an elementary conclusion is that the smallest perturbation E to a matrix A that is needed to make a vector x satisfy $(A+E)x = b$ is

$$\mu^{(\mathrm{LE})}(x) \;=\; \min_{E \,:\, (A+E)x \,=\, b} \|E\| \;=\; \frac{\|Ax - b\|}{\|x\|}\,,$$

where the matrix has the operator norm determined by the two vector norms. (An explanation of the specialized norms that are sometimes used in this and other cases is beyond the scope of this discussion.) Formulas for various minima were found subsequently for perturbations to the data of symmetric linear equations [7] [50], of Toeplitz equations [19] [62], of Vandermonde equations [3] [58], of equations with multiple right sides [20], and of minimum-norm solutions for underdetermined linear equations [59].

Least squares: Waldén, Karlson and Sun [63] were the first to find formulas for the minimal size of backward errors for least squares problems (LS). For example of their results, consider the the linear regression problem,

$$\min_v \|b - Av\|_2\,,$$

which is equivalent to the equation $A^t(b - Av) = 0$. The size of smallest perturbations E to the matrix A that is needed to make a given vector x a solution of the linear regression problem is

$$\mu^{(\mathrm{LS})}(x) \;=\; \min_{E \,:\, (A+E)^t[b-(A+E)x] \,=\, 0} \|E\|_F^2 \;=\; \frac{\|r\|_2^2}{\|x\|_2^2} + \min\{0, \lambda\}\,,$$

where $\|\cdot\|_F$ is the Frobenius norm, where $r = b - Ax$, and where λ is the smallest eigenvalue of a certain matrix,

$$\lambda = \lambda_{\min}\left(AA^t - \frac{rr^t}{\|x\|_2^2}\right)\,.$$

An alternate expression for the minimum was given by Higham [22, p. 405] [63, p. 275]. Other estimates and formulas were found for data perturbations for these problems [17] [25] [30] [57], for problems with with multiple right sides [56], for linearly constrained problems [10], and for spherically constrained least squares problems [29].

Matrix factorizations: Minimal backward errors for matrix factorizations have been considered for the Choleski and QR factorizations [53], and for spectral decompositions of Hermitian matrices [8] [55].

Invariant subspaces: Closely related to factorization of matrices is the construction of invariant subspaces of linear operators. Minimal backward errors were considered for individual eigenvalues and eigenvectors [13] [21] [51], for Krylov subspaces [52], and for other invariant subspaces [54].

2. Equivalent Optimizations

2.1. **Overview of the Approach.** The approach used here to study perturbed metric projections is a formalization of one introduced by Shapiro [44] [46]. The idea is to study the local response of a parameterized optimization problem by simplifying the problem in a way that does not materially alter the response.

Specifically, for both of the applications discussed in the Introduction, the value

$$\mu(x) = \min_{y \,:\, F(y,x) \,=\, 0} \|y - y_0\| \tag{2}$$

has interest only when $x \approx x_0$ where (y_0, x_0) solves the constraint equation, $F(y, x) = 0$. As the parameter x approaches x_0, the implicit function theorem implies that the points where the minima occur converge to y_0, in which case the constraint function may be approximated by its tangent at the limit (assuming the function is sufficiently smooth). In this way one is naturally led to the idea of perturbing equation (2) by altering the constraint. To that end, there are two basic requirements for the altered problems.

- The optimal values of the altered problems should mimic how $\mu(x)$ varies with x.
- Since $\mu(x)$ is of interest only when $x \approx x_0$, good approximations are needed only near x_0.

The novelty of the present approach is to formalize these requirements by an equivalence relation among functions of x. The relation is chosen so that if two functions are equivalent, then it shall be agreed that they are acceptable estimates for one another. Next, parameterized minimization problems are identified that are simpler than equation (2) but whose optimal values belong to the same equivalence class as $\mu(x)$. In this way equation (2) is made simpler while changing the optimal-value function only in acceptable ways. For the purposes of this paper, "acceptable" means that $\mu(x)$'s first-order sensitivity to the parameter x, at x_0, should be invariant with respect to the changes in the function.

2.2. **Equivalence Relations.** The following equivalence relation is appropriate when differentiability at x_0 is the object of study.

Definition 2.1 (Differential Equivalence)**.** The functions f and g defined on a neighborhood of $x_0 \in \mathbb{R}^n$ with values in $\mathbb{R}^p$ are called differentially equivalent,

$$f \underset{x_0}{\overset{\partial}{\simeq}} g\,,$$

provided $f - g$ is Fréchet differentiable at x_0 and the derivative vanishes there; equivalently,

$$\lim_{x \,\to\, x_0} \frac{\|f(x) - g(x)\|}{\|x - x_0\|} \;=\; 0\,. \tag{3}$$

Lemma 2.2. *Definition 2.1's* $\simeq^{\partial}_{x_0}$ *is an equivalence relation.*

If g is an affine function, then equation (3) is essentially the definition for the Fréchet derivative of f at x_0. If additionally the limit is restricted to a ray emanating from x_0, then equation (3) implies a form of directional derivative. In this way f's differential properties are determined by its equivalence class.

A simpler equivalence relation is that any approximation to $f(x)$ should be relatively more accurate as x approaches x_0.

Definition 2.3 (Rational Equivalence)**.** The real-valued functions f and g defined on a neighborhood of $x_0 \in \mathbb{R}^n$ are called rationally equivalent,

$$f \underset{x_0}{\overset{\div}{\simeq}} g\,,$$

provided for every $\epsilon > 0$ there is a neighborhood $N(\epsilon)$ of x_0 such that $x \in N(\epsilon)$ implies

$$(1 - \epsilon) g(x) \le f(x) \le (1 + \epsilon) g(x) \, . \tag{4}$$

Lemma 2.4. *Definition 2.3's* $\simeq^{\div}_{x_0}$ *is an equivalence relation.*

Definition 2.3's rational equivalence is stronger than Definition 2.1's differential equivalence. For example, two monomials $c_1 x^{n_1}$ and $c_2 x^{n_2}$ are rationally equivalent at 0 if an only if they are equal, but all monomials with vanishing derivatives at 0 are differentially equivalent there.

For equation (2)'s function $\mu(x)$, Definition 2.3 implies Definition 2.1. Thus equation (4), which is easier to verify in proofs, imposes equation (3)'s differential approximation. The proof of this is based on the Lipschitz continuity of $\mu(x)$ at x_0, which is easily established for metric projections. The following notation and assumptions are used in this proof and throughout the paper.

Hypothesis 2.5. Assume

- $\mathcal{D} \subseteq \mathbb{R}^m \times \mathbb{R}^n$ is a neighborhood of (y_0, x_0),
- $F : \mathcal{D} \to \mathbb{R}^p$ is continuously differentiable.

Notation 2.6. Under Hypotheses 2.5,

- When it exists,
$$\mu(x) = \min_{y \, : \, F(y,x) \, = \, 0} \| y - y_0 \| \, .$$
- $D_1 F(y, x) \in \hom(\mathbb{R}^m, \mathbb{R}^p)$ is the partial Fréchet derivative of F at (y, x) with respect to the first block of variables, and similarly for $D_2 F$ with respect to the second block of variables.
- $\mathcal{T}_{y_0}(y, x) = D_1 F(y_0, x)(y - y_0) + F(y_0, x)$ is the linear function of y, parameterized by x, whose graph is tangent to $F(y, x)$'s at $y = y_0$.

There seems to be no standard notation for the partial derivatives in *Notation* 2.6: "$D_{(1)} F$" is used in [4], "$\partial_1 F$" in [36], and "$D_y F$" in [6]. This paper chooses to write "$D_1 F$."

Lemma 2.7 (Lipschitz Continuity of $\mu(x)$ at x_0)**.**

- *In addition to Hypothesis 2.5, suppose that* $F(y_0, x_0) = 0$*, and that* $D_1 F(y_0, x_0)$ *is onto.*

$\Rightarrow$ *There is a constant* $L > 0$ *and a neighborhood* $N^{(2.7)}_{x_0}$ *of* x_0 *where Definition 2.6's function* $\mu(x)$ *exists, and* $\mu(x) \le L \| x - x_0 \|$*.*

Proof. The implicit function theorem says x_0 has a neighborhood N on which there is a continuously differentiable function $\phi : N \to \mathbb{R}^m$ such that $\phi(x_0) = y_0$ and $F(\phi(x), x) = 0$. Thus $\mu(x)$'s minimization problems have feasible points for all $x \in N$. The feasible sets are closed because F is continuous, so the minimal distance to y_0 is attained. This means μ is well-defined on N. Since ϕ is continuously differentiable, x_0 has a compact, convex neighborhood $N^{(2.7)}_{x_0} \subseteq N$ on which ϕ is Lipschitz continuous with Lipschitz constant L. Thus $\mu(x) \le \| \phi(x) - y_0 \| = \| \phi(x) - \phi(x_0) \| \le L \| x - x_0 \|$ for every $x \in N^{(2.7)}_{x_0}$. □

Corollary 2.8. *Continuing Lemma 2.7, for Notation 2.6's function $\mu(x)$ and for any function f,*

$$f \underset{x_0}{\overset{\div}{\simeq}} \mu \quad \Rightarrow \quad f \underset{x_0}{\overset{\partial}{\simeq}} \mu \, .$$

Proof. Let $N(\epsilon)$ be the neighborhoods in Definition 2.3 for the equivalence $f \simeq^{\div}_{x_0} \mu$, and let $N^{(2.7)}_{x_0}$ be Lemma 2.7's neighborhood. If $x \in N(\epsilon) \cap N^{(2.7)}_{x_0}$, then $(1 - \epsilon)\mu(x) \le f(x) \le (1 + \epsilon)\mu(x)$, so $|f(x) - \mu(x)| \le \epsilon\, \mu(x) \le \epsilon L \|x - x_0\|$, which proves that the limit in equation (3) vanishes. □

2.3. **Equivalence Classes.** All the optimization problems in Table 4 are differentially equivalent at x_0, in the sense of Definition 2.1. The Table's problems differ from the original problem, (P), first by linearizing the constraint with respect to the dependent variable y, which is problem $(P_{\mathcal{T}})$, and then by additionally linearizing the constraint with respect to the independent parameter x, which is problem (P_ℓ). An intermediate problem, (P_0), removes the bilinear variation in problem $(P_{\mathcal{T}})$'s constraint by fixing the bilinear term at x_0.

Table 4. Optimization problems parameterized by x whose optimal values are differentially equivalent to *Notation* 2.6's function $\mu(x)$ at $x = x_0$. Besides Hypothesis 2.5, it is assumed that $F(y_0, x_0) = 0$, and $D_1 F(y_0, x_0)$ is onto. These expressions use the notation $\Delta x = x - x_0$ and $\Delta y = y - y_0$.

name	value	minimization form	dual, maximization form
(P)	$\mu(x)$	$\min\limits_{y\,:\,F(y,x)=0} \quad \|\Delta y\|$	
$(P_{\mathcal{T}})$	$\mu_{\mathcal{T}}(x)$	$\min\limits_{y\,:\,D_1 F(y_0,x)\Delta y + F(y_0,x) = 0} \quad \|\Delta y\|$	$\max\limits_{f\,:\,\|D_1 F(y_0,x)^* f\| \le 1} \quad f(F(y_0,x))$
(P_0)	$\mu_0(x)$	$\min\limits_{y\,:\,D_1 F(y_0,x_0)\Delta y + F(y_0,x) = 0} \quad \|\Delta y\|$	$\max\limits_{f\,:\,\|D_1 F(y_0,x_0)^* f\| \le 1} \quad f(F(y_0,x))$
(P_ℓ)	$\mu_\ell(x)$	$\min\limits_{y\,:\,DF(y_0,x_0)(\Delta y, \Delta x) = 0} \quad \|\Delta y\|$	$\max\limits_{f\,:\,\|D_1 F(y_0,x_0)^* f\| \le 1} \quad f(D_2 F(y_0,x_0)\Delta x)$

What is remarkable about the present situation is that Definition 2.1's equivalence class for $\mu(x)$ is invariant with respect to all these changes. Whether this is true for parameterized optimization problems that are more general than metric projections with equality constraints, is open. Figure 1 indicates where Table 4's equivalences are established in this paper.

3. First Equivalence, $(P)_{\min} \Leftrightarrow (P_{\mathcal{T}})_{\min}$

The first equivalence to be proved (in the notation of Table 4),

$$\mu(x)_{\min} \underset{x_0}{\overset{\div}{\simeq}} \mu_{\mathcal{T}}(x)_{\min} \, ,$$

says that the nonlinear constraint $F(y, x) = 0$ can be replaced by the linear constraint $D_1 F(y_0, x)\Delta y + F(y_0, x) = 0$. Figure 2 depicts the several lemmas that contribute to the proof of this. Those in the central column — Lemma 3.1, Corollary 3.2, and Lemma 3.10 — are interesting in their own right. The first two extend

$$\begin{tikzcd}
(P)_{\min} \arrow[d, Leftrightarrow, "\text{Thm. 3.11}"', "\textit{rational equivalence}"] & \\
(P_{\mathcal{T}})_{\min} \arrow[r, Leftrightarrow, "\text{Thm. 4.4}", "\textit{duality equality}"'] & (P_{\mathcal{T}})_{\max} \arrow[d, Leftrightarrow, "\text{Thm. 5.1}"', "\textit{rational equivalence}"] \\
(P_0)_{\min} \arrow[r, Leftrightarrow, "\text{Thm. 4.4}", "\textit{duality equality}"'] & (P_0)_{\max} \arrow[d, Leftrightarrow, "\text{Thm. 6.4}"', "\textit{differential equivalence}"] \\
(P_\ell)_{\min} \arrow[r, Leftrightarrow, "\text{Thm. 4.4}", "\textit{duality equality}"'] & (P_\ell)_{\max}
\end{tikzcd}$$

Figure 1. Where and how Table 4's equivalences are proved.

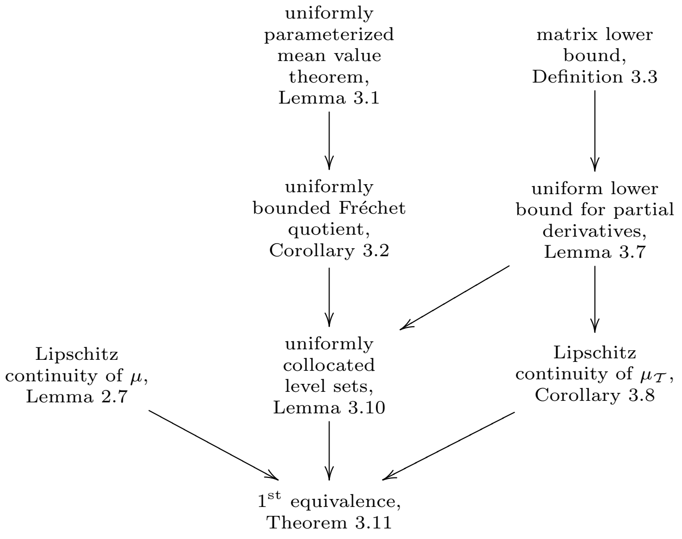
uniformly parameterized mean value theorem, Lemma 3.1
matrix lower bound, Definition 3.3
uniformly bounded Fréchet quotient, Corollary 3.2
uniform lower bound for partial derivatives, Lemma 3.7
Lipschitz continuity of μ, Lemma 2.7
uniformly collocated level sets, Lemma 3.10
Lipschitz continuity of $\mu_{\mathcal{T}}$, Corollary 3.8
1^{st} equivalence, Theorem 3.11

Figure 2. Dependencies for the proof of the first equivalence.

some basic results in real analysis to parameterized functions. The third is related to the implicit function theorem in that it provides a local description of a function's level sets.

3.1. **Uniformly Parameterized Mean Value Theorem.**

It is well known that if f is continuously differentiable, then for every y_0 and every $\epsilon > 0$ there is a neighborhood $N_{y_0}(\epsilon)$ of y_0 where

$$y_1, y_2 \in N_{y_0}(\epsilon) \;\Rightarrow\; \|f(y_1) - f(y_2) - Df(y_0)(y_1 - y_2)\| \le \epsilon \, \|y_1 - y_2\| \,. \tag{5}$$

This serves as a mean value theorem in multiple dimensions. It has been discussed many times [28, p. 212, notes for §7.1–4]. For example, Bartle [4, p. 377, lemma 41.4] calls it the "key lemma" for mapping theorems that are related to the implicit function theorem. Ortega and Rheinboldt [36, p. 72, lemma 3.2.10] show that equation (5) actually is equivalent to the continuity of the derivative. Here, this surrogate mean value theorem is generalized to parameterized functions.

Lemma 3.1 (Uniformly Parameterized Mean Value Theorem)**.**

- *Assume Hypothesis 2.5.*

⇒ *For every $\epsilon > 0$ there is a neighborhood $N^{(3.1)}_{y_0}(\epsilon) \times N^{(3.1)}_{x_0}(\epsilon) \subseteq \mathcal{D}$ of (y_0, x_0) such that for all $y_1, y_2, y_3 \in N^{(3.1)}_{y_0}(\epsilon)$ and all $x \in N^{(3.1)}_{x_0}(\epsilon)$,*

$$\|F(y_1, x) - F(y_2, x) - D_1 F(y_3, x)(y_1 - y_2)\| \le \epsilon \, \|y_1 - y_2\| \,. \tag{6}$$

Proof. It is well known [4, p. 376, lemma 41.3] [36, p. 70, lemma 3.2.5] that if $D \subseteq \mathbb{R}^m$ is a convex, open set, and if $f : D \to \mathbb{R}^p$ is continuously differentiable, then for any $y_1, y_2, y_3 \in D$,

$$\begin{aligned} &\|f(y_1) - f(y_2) - Df(y_3)(y_1 - y_2)\| \\ &\qquad \le \sup_{0 \,\le\, t \,\le\, 1} \|Df(ty_1 + (1-t)y_2) - Df(y_3)\| \, \|y_1 - y_2\| \,. \end{aligned}$$

It is always possible to find a convex neighborhood Y_0 of y_0, and a neighborhood X_0 of x_0, so that $Y_0 \times X_0 \subseteq \mathcal{D}$. Thus, for any any $y_1, y_2, y_3 \in Y_0$ and $x \in X_0$,

$$\begin{aligned} &\|F(y_1, x) - F(y_2, x) - D_1 F(y_3, x)(y_1 - y_2)\| \\ &\qquad \le \sup_{0 \,\le\, t \,\le\, 1} \|D_1 F(ty_1 + (1-t)y_2, x) - D_1 F(y_3, x)\| \, \|y_1 - y_2\| \,. \end{aligned} \tag{7}$$

It is further possible to choose Y_0 and X_0 so that $Y_0 \times X_0$ is bounded and $\mathrm{cl}(Y_0 \times X_0) \subseteq \mathcal{D}$. Thus $K = \mathrm{cl}(Y_0) \times \mathrm{cl}(Y_0) \times \mathrm{cl}(X_0)$ is compact. Since $D_1 F(y, x)$ is continuous, therefore $g(y_1, y_2, x) = D_1 F(y_1, x) - D_1 F(y_2, x)$ is uniformly continuous on K. Let the norm for the domain of $g(y_1, y_2, x)$ be $\max\{\|y_1\|, \|y_2\|, \|x\|\}$. The uniform continuity means, for every $\epsilon > 0$ there is a $\delta(\epsilon) > 0$ so that if $(y_1, y_2, x), (y_1', y_2', x') \in K$ with $\max\{\|y_1 - y_1'\|, \|y_2 - y_2'\|, \|x - x'\|\} \le \delta(\epsilon)$, then $\|g(y_1, y_2, x) - g(y_1', y_2', x')\| \le \epsilon$.

If y_1, $y_2 \in \mathcal{B}_{y_0}(\delta(\epsilon)) \cap Y_0$, then all convex combinations of these points also lie in this set. If additionally $y_3 \in \mathcal{B}_{y_0}(\delta(\epsilon)) \cap Y_0$ and $x \in \mathcal{B}_{x_0}(\delta(\epsilon)) \cap X_0$, then

$$\begin{aligned} &\|D_1 F(ty_1 + (1-t)y_2, x) - D_1 F(y_3, x)\| \\ &\quad = \|g(ty_1 + (1-t)y_2, y_3, x)\| \\ &\quad = \|g(ty_1 + (1-t)y_2, y_3, x) - g(y_0, y_0, x_0)\| \\ &\quad \le \epsilon \,. \end{aligned}$$

Thus the supremum in equation (7) is bounded by ϵ provided $y_1, y_2, y_3 \in \mathcal{B}_{y_0}(\delta(\epsilon)) \cap Y_0$ and $x \in \mathcal{B}_{x_0}(\delta(\epsilon)) \cap X_0$. □

Lemma 3.1 gives conditions under which the Fréchet differential is uniformly approximating with respect to a parameter.

Corollary 3.2 (Uniformly Bounded Fréchet Quotient)**.** *Lemma 3.1's neighborhoods also satisfy, for all $y \in N^{(3.1)}_{y_0}(\epsilon)$ and $x \in N^{(3.1)}_{x_0}(\epsilon)$,*

$$\|F(y,x) - \mathcal{T}_{y_0}(y,x)\| \le \epsilon \, \|y - y_0\| \, , \tag{8}$$

where $\mathcal{T}_{y_0}(y,x)$ is the parameterized tangent function of Notation 2.6.

Proof. Choose $y_1 = y$, $y_2 = y_0$ and $y_3 = y_0$ so that the formula in Lemma 3.1's equation (6) becomes

$$\begin{aligned} &F(y_1,x) - F(y_2,x) - D_1F(y_3,x)(y_1 - y_2) \\ &\quad = F(y,x) - F(y_0,x) - D_1F(y_0,x)(y - y_0) \\ &\quad = F(y,x) - \mathcal{T}_{y_0}(y,x)\, . \end{aligned}$$

□

3.2. **Matrix Lower Bound.** The matrix lower bound, $\|A\|_\ell$, is analogous to the matrix norm but with subtly different properties. The concept was first described by von Neumann [34, p. 1042] essentially for nonsingular matrices. It can be extended to all nonzero matrices using ideas of Banach [1, p. 150, chapter 10, theorem 10]. See [16] for further discussion of this concept.

Definition 3.3 (Matrix Lower Bound [16, def. 2.1])**.** Let A be a nonzero matrix. The matrix lower bound, $\|A\|_\ell$, is the largest of the numbers, m, such that for every y in the column space of A, there is some x with $Ax = y$ and $m\,\|x\| \le \|y\|$.

Lemma 3.4. [16, lem. 2.2] *The matrix lower bound exists and is positive for every nonzero matrix.*

Lemma 3.5. [16, cor. 4.3] *The matrix lower bound is continuous on the set of full rank matrices.*

Lemma 3.6. [16, thm. 5.1] *The matrix lower bound of a full rank matrix is the distance to the set of rank deficient matrices.*

The immediate use of the lower bound is the following lemma, which is used to show that $\mu_{\mathcal{T}}(x)_{\min}$ is well-defined in a neighborhood of x_0 and Lipschitz continuous at x_0.

Lemma 3.7 (Uniform Lower Bounds for Partial Derivatives)**.**

- *In addition to Hypothesis 2.5, suppose that $D_1F(y_0,x_0)$ is onto.*

⇒ *There is a compact neighborhood $K^{(3.7)}_{x_0}$ of x_0 and a number $m_\ell > 0$ such that every $x \in K^{(3.7)}_{x_0}$ and $u \in \mathbb{R}^p$ have some $w \in \mathbb{R}^m$ (which depends on x and u) so that $D_1F(y_0,x)w = u$ and $m_\ell\|w\| \le \|u\|$.*

Proof. Choose some bases for $\mathbb{R}^m$ and $\mathbb{R}^p$ so that these spaces are represented by real column vectors. The linear transformations $D_1F(y_0,x)$ are then represented by $p \times m$ matrices, $A(x)$. These matrices are continuous functions of x because by hypothesis F is continuously differentiable. That $D_1F(y_0,x_0)$ is onto is equivalent to $A(x_0)$ having full row rank, p. From Lemmas 3.4 and 3.6, $A(x_0)$ has a neighborhood of matrices $N_{A(x_0)}$ all of which have full row rank. From the continuity of $A(x)$, x_0 has a neighborhood N_{x_0} all of whose matrices lie in $N_{A(x_0)}$. Hence the mappings $D_1F(y_0,x)$ are onto for all $x \in N_{x_0}$. Finally, since $\|A(x)\|_\ell$ is continuous and positive on N_{x_0} by Lemmas 3.4 and 3.5, it is possible to choose a compact neighborhood $K^{(3.7)}_{x_0} \subseteq N_{x_0}$ of x_0 where $\|A(x)\|_\ell$ is uniformly bounded below by some $m_\ell > 0$. □

Corollary 3.8 (Lipschitz Continuity of $\mu_{\mathcal{T}}(x)_{\min}$ at x_0)**.**

- *In addition to Hypothesis 2.5, suppose that $F(y_0, x_0) = 0$, and that $D_1 F(y_0, x_0)$ is onto.*

⇒ *There is a constant $L_{\mathcal{T}} > 0$ and a neighborhood $N^{(3.8)}_{x_0}$ of x_0 where Table 4's function $\mu_{\mathcal{T}}(x)_{\min}$ is well-defined and $\mu_{\mathcal{T}}(x)_{\min} \le L_{\mathcal{T}} \|x - x_0\|$.*

Proof. Let $K^{(3.7)}_{x_0}$ be Lemma 3.7's compact neighborhood for x_0. If $x \in K^{(3.7)}_{x_0}$, then by Lemma 3.7 there is a $y \in \mathbb{R}^m$ such that

$$\mathcal{T}_{y_0}(y, x) = D_1 F(y_0, x)(y - y_0) + F(y_0, x) = 0$$

and $m_\ell \|y - y_0\| \le \|F(y_0, x)\|$. The equality means that the minimization for $\mu_{\mathcal{T}}(x)_{\min}$ has a feasible point. Since $\mathcal{T}_{y_0}$ is continuous, the feasible set is closed, so the minimum is attained. Thus $\mu_{\mathcal{T}}(x)_{\min}$ is well-defined on $K^{(3.7)}_{x_0}$. It is always possible to find a convex subneighborhood $N^{(3.8)}_{x_0} \subseteq K^{(3.7)}_{x_0}$. Since $F(y_0, x)$ is continuously differentiable, it is also Lipschitz continuous on this set. Let L be the Lipschitz constant. Finally, for $x \in N^{(3.8)}_{x_0}$,

$$\mu_{\mathcal{T}}(x)_{\min} \le \|y - y_0\| \le \frac{\|F(y_0, x)\|}{m_\ell} = \frac{\|F(y_0, x) - F(y_0, x_0)\|}{m_\ell} \le \frac{L\|x - x_0\|}{m_\ell}\,.$$

□

3.3. **Uniformly Collocated Level Sets.** Suppose D is an open set in $\mathbb{R}^m$, on which $f : D \to \mathbb{R}^n$ is continuously differentiable. By analogy with real-valued functions, the set $f^{-1}(y)$ may be called a level set of f.

If $f(x_0) = 0$ and the linear transformation $Df(x_0) : \mathbb{R}^m \to \mathbb{R}^n$ is onto, the implicit function theorem says the level set $f^{-1}(0)$ contains the graph of a smooth curve passing through x_0. Usually the implicit function (which parameterizes the curve) is emphasized, but the theorem also can be interpreted as describing the level set of roots. For example [4, p. 384, theorem 41.9 part (b)], there is neighborhood N_{x_0} of x_0 where the implicit function's graph is the entire level set.

It is possible to make a geometric comparison between all the level sets of f and those of its tangent function at x_0. Near x_0, the corresponding level sets are always present and they are asymptotically identical [16]. The proof of this is a modification of a construction apparently due to L. M. Graves [15], see also [4, p. 378, theorem 41.6]. Here, the collocation result is extended to functions that vary smoothly with a parameter. In this case the distance between the corresponding level sets is uniformly bounded with respect to changes in the parameter.

Notation 3.9. In whatever space is indicated, let $\mathcal{B}_c(r)$ be the open ball with center c and radius r.

Lemma 3.10 (Uniformly Collocated Level Sets)**.**

- *In addition to Hypothesis 2.5, suppose that $D_1 F(y_0, x_0)$ is onto.*
- *Let $\mathcal{T}_{y_0}(y, x)$ be Definition 2.6's linear function of y, parameterized by x, whose graph is tangent to $F(y, x)$'s at $y = y_0$.*

⇒ *For every $\epsilon > 0$ there is an $r(\epsilon) > 0$ and a neighborhood $N^{(3.10)}_{x_0}(\epsilon)$ of x_0 so $\mathcal{B}_{y_0}(r(\epsilon)) \times N^{(3.10)}_{x_0}(\epsilon) \subseteq \mathcal{D}$.*

$\Rightarrow$ *For each pair* $(y, x) \in \mathcal{B}_{y_0}(r(\epsilon)/(1+\epsilon)) \times N_{x_0}^{(3.10)}(\epsilon)$,

(1) *there exists* $y_{\mathcal{T}} \in \mathcal{B}_{y_0}(r(\epsilon))$ *with*

$$\|y_{\mathcal{T}} - y\| \le \epsilon \, \|y - y_0\| \text{ and } \mathcal{T}_{y_0}(y_{\mathcal{T}}, x) = F(y, x), \tag{9}$$

(2) *and there exists* $y_F \in \mathcal{B}_{y_0}(r(\epsilon))$ *with*

$$\|y_F - y\| \le \epsilon \, \|y - y_0\| \text{ and } \mathcal{T}_{y_0}(y, x) = F(y_F, x).$$

Proof. The proof is based on Lemma 3.1's neighborhoods $N_{y_0}^{(3.1)}(\rho)$ and $N_{x_0}^{(3.1)}(\rho)$, for a ρ determined from ϵ as follows. Let $\delta = \epsilon/(1+\epsilon) < 1$. Let m_ℓ and $K_{x_0}^{(3.7)}$ be as in Lemma 3.7. It is always possible to find an $r(\epsilon) > 0$ so that

$$\mathrm{cl}\,(\mathcal{B}_{y_0}(r(\epsilon))) \subseteq N_{y_0}^{(3.1)}(\delta \, m_\ell)\,. \tag{10}$$

With this preparation, the neighborhoods from which the theorem is allowed to choose y and x are then

$$y \in \mathcal{B}_{y_0}(r(\epsilon)/(1+\epsilon)) \subseteq \mathrm{cl}\,(\mathcal{B}_{y_0}(r(\epsilon))) \;\subseteq\; N_{y_0}^{(3.1)}(\delta \, m_\ell)\,, \tag{11}$$

$$x \in N_{x_0}^{(3.10)}(\epsilon) := K_{x_0}^{(3.7)} \cap N_{x_0}^{(3.1)}(\delta \, m_\ell) \;\subseteq\; N_{x_0}^{(3.1)}(\delta \, m_\ell)\,. \tag{12}$$

Note that $\mathcal{B}_{y_0}(r(\epsilon)) \times N_{x_0}^{(3.10)}(\epsilon) \subseteq N_{y_0}^{(3.1)}(\delta \, m_\ell) \times N_{x_0}^{(3.1)}(\delta \, m_\ell) \subseteq \mathcal{D}$ as required.

(Part 1.) Since $x \in K_{x_0}^{(3.7)}$, there is a $y_{\mathcal{T}}$ with

$$\begin{aligned} D_1 F(y_0, x)(y_{\mathcal{T}} - y) &= F(y, x) - \mathcal{T}_{y_0}(y, x)\,, \\ m_\ell \, \|y_{\mathcal{T}} - y\| &\le \|F(y, x) - \mathcal{T}_{y_0}(y, x)\|\,. \end{aligned}$$

The equality and some algebra imply $\mathcal{T}_{y_0}(y_{\mathcal{T}}, x) = F(y, x)$, which is part of equation (9), while the inequality and equations (11) and (8) imply

$$\|y_{\mathcal{T}} - y\| \le \frac{\|F(y, x) - \mathcal{T}_{y_0}(y, x)\|}{m_\ell} \le \frac{\delta \, m_\ell \, \|y - y_0\|}{m_\ell} = \delta \|y - y_0\| < \epsilon \|y - y_0\|\,,$$

which is the other part of equation (9). From this follows $\|y_{\mathcal{T}} - y_0\| \le (1+\epsilon)\|y - y_0\|$, so $y_{\mathcal{T}} \in \mathcal{B}_{y_0}(r(\epsilon))$ by the choice of y in equation (11). This membership is the remaining conclusion in Part 1.

(Part 2.) Let $y_1 = y$, so the following conditions are satisfied for $j = 0$,

$$\begin{aligned} &(1_j) \quad \|y_{j+1} - y_j\| \le \delta^j \, \|y - y_0\|\,, \\ &(2_j) \quad \|F(y_{j+1}, x) - \mathcal{T}_{y_0}(y, x)\| \le \delta \, m_\ell \, \|y_{j+1} - y_j\|\,. \end{aligned}$$

The first is trivial; the second is by Corollary 3.2 (with $\epsilon = \delta \, m_\ell$), which is applicable by the choice of y and x in equations (11) and (12).

Summing (1_j) for $0 \le j \le k$ gives

$$\|y_{k+1} - y_0\| \le \sum_{j=0}^{k} \|y_{j+1} - y_j\| \le \frac{1 - \delta^{k+1}}{1 - \delta} \, \|y - y_0\| < (1+\epsilon) \, \|y - y_0\|\,.$$

This combines with $y \in \mathcal{B}_{y_0}(r(\epsilon)/(1+\epsilon))$ to place $y_{k+1} \in \mathcal{B}_{y_0}(r(\epsilon))$. Therefore $y_{k+1} \in N_{y_0}^{(3.1)}(\delta \, m_\ell)$ from equation (10), so the evaluation of $F(y_{k+1}, x)$ in condition (2_k) is always well-defined when (1_j) holds for $0 \le j \le k$.

Suppose y_0, y_1, …, y_n have been constructed to satisfy (1_j) and (2_j) for $0 \le j \le n-1$. As in Part (1) but with a change of sign, Lemma 3.7 says there is a y_{n+1} with

$$\begin{aligned} D_1 F(y_0, x)(y_{n+1} - y_n) &= \,- \left[F(y_n, x) - \mathcal{T}_{y_0}(y_1, x)\right], \\ \|D_1 F(y_0, x)\|_\ell \, \|y_{n+1} - y_n\| &\le \|F(y_n, x) - \mathcal{T}_{y_0}(y, x)\| \,. \end{aligned}$$

The inequality and conditions (2_{n-1}) and (1_{n-1}) imply condition (1_n),

$$\begin{aligned} \|y_{n+1} - y_n\| \; &\le \; \frac{\|F(y_n, x) - \mathcal{T}_{y_0}(y, x)\|}{\|D_1 F(y_0, x)\|_\ell} \\ &\le \; \frac{\delta \, m_\ell \, \|y_n - y_{n-1}\|}{m_\ell} \\ &\le \; \frac{\delta \, m_\ell \, \delta^{n-1} \|y - y_0\|}{m_\ell} \\ &= \; \delta^n \, \|y - y_0\| \,. \end{aligned}$$

It is therefore possible to evaluate $F(y_{n+1}, x)$. Condition (2_n) now holds since

$$\begin{aligned} \|F(y_{n+1}, x) - \mathcal{T}_{y_0}(y, x)\| \; &= \; \|F(y_{n+1}, x) - F(y_n, x) - D_1 F(y_0, x)(y_{n+1} - y_n)\| \\ &\le \; \delta \, m_\ell \|y_{n+1} - y_n\| \,. \end{aligned}$$

The equality above is from the choice of y_{n+1} and from some algebra which, beware, is not straightforward; the inequality is from equation (6), which is applicable because y_n, $y_{n+1} \subseteq N^{(3.1)}_{y_0}(\delta \, m_\ell)$.

In this way a sequence $\{y_n\} \subseteq \mathcal{B}_{y_0}(r(\epsilon))$ is constructed that satisfies conditions (1_n) and (2_n) for all n. This sequence establishes the three conclusions of Part 2. The sequence is a Cauchy sequence by (1_n), so it has a limit $y_F \in \mathrm{cl}(\mathcal{B}_{y_0}(r(\epsilon)))$. Passing to the limit in (2_n) shows $F(y_F, x) = \mathcal{T}_{y_0}(y, x)$. Summing (1_j), now for $1 \le j \le n$, gives

$$\|y_{n+1} - y\| \le \|y_{n+1} - y_1\| \le \sum_{j=1}^{n} \|y_{j+1} - y_j\| \le \delta \, \frac{1 - \delta^n}{1 - \delta} \, \|y - y_0\| \,,$$

which in the limit becomes $\|y_F - y\| \le \delta(1-\delta)^{-1}\|y - y_0\| = \epsilon \|y - y_0\|$. □

3.4. **Proof of the First Equivalence.**

Theorem 3.11 ($(P)_{\min} \Leftrightarrow (P_{\mathcal{T}})_{\min}$)**.**

- *In addition to Hypothesis 2.5, suppose that* $F(y_0, x_0) = 0$*, and that* $D_1 F(y_0, x_0)$ *is onto.*

⇒ *There is a neighborhood of* x_0 *where both of Table 4's optimization problems* $(P)_{\min}$ *and* $(P_{\mathcal{T}})_{\min}$ *are well-defined. Their values are rationally equivalent at* x_0 *in the sense of Definition 2.3.*

Proof. By Lemma 2.7, x_0 has a neighborhood $N^{(2.7)}_{x_0}$ where for every $x \in N^{(2.7)}_{x_0}$ problem $(P)_{\min}$ is well-defined, and the optimal value, $\mu(x)$, is Lipschitz continuous at x_0 with constant L.

Similarly, by Corollary 3.8, x_0 has a neighborhood $N^{(3.8)}_{x_0}$ where problem $(P_{\mathcal{T}})_{\min}$ is well-defined for every $x \in N^{(3.8)}_{x_0}$, and the optimal value, $\mu_{\mathcal{T}}(x)$, is Lipschitz continuous at x_0 with constant $L_{\mathcal{T}}$.

Finally, let $\mathcal{B}_{y_0}(r(\epsilon)/(1+\epsilon)) \times N_{x_0}^{(3.10)}(\epsilon)$ be Lemma 3.10's neighborhood of (y_0, x_0), and let

$$N(\epsilon) = N_{x_0}^{(2.7)} \cap N_{x_0}^{(3.8)} \cap N_{x_0}^{(3.10)}(\epsilon) \cap \mathcal{B}_{x_0}\left(\frac{r(\epsilon)}{1+\epsilon} \min\{L^{-1}, L_{\mathcal{T}}^{-1}\}\right).$$

Suppose $x \in N(\epsilon)$. Let $\mu(x)$ be attained at y. By Lemma 2.7 and since $x \in \mathcal{B}_{x_0}(L^{-1} r(\epsilon)/(1+\epsilon))$, therefore

$$\|y - y_0\| = \mu(x) \le L\|x - x_0\| < r(\epsilon)/(1+\epsilon),$$

which places $(y, x) \in \mathcal{B}_{y_0}(r(\epsilon)/(1+\epsilon)) \times N_{x_0}^{(3.10)}(\epsilon)$. Part 1 of Lemma 3.10 now asserts there is a $y_{\mathcal{T}} \in \mathcal{B}_{y_0}(r(\epsilon))$ with

$$\|y_{\mathcal{T}} - y\| \le \epsilon\, \|y - y_0\| \quad \text{and} \quad \mathcal{T}_{y_0}(y_{\mathcal{T}}, x) = F(y, x) = 0\,.$$

Thus

$$\mu_{\mathcal{T}}(x) \le \|y_{\mathcal{T}} - y_0\| \le \|y_{\mathcal{T}} - y\| + \|y - y_0\| \le (1+\epsilon)\, \|y - y_0\| = (1+\epsilon)\, \mu(x)$$

which is half of Definition 2.3's inequality. The same proof establishes the inequality with μ and $\mu_{\mathcal{T}}$ exchanged, using Corollary 3.8 instead of Lemma 2.7, and Lemma 3.10 part 2 instead of part 1. □

4. Equalities for the Dual Problems

The duality theory for best linear approximation guarantees that the three pairs of dual problems in Table 4 have equal values. Equalities like these are well known and can be established in many ways. These are derived from a duality theorem that Luenberger [28] proves directly from the Hahn-Banach theorem.

Theorem 4.1 (Best Linear Approximation [28, p. 119, theorem 1])**.** *If $\mathcal{S}$ is a subspace and y_0 is an element of a real, normed linear space, then*

$$\inf_{y \,\in\, \mathcal{S}} \|y - y_0\| \;=\; \max_{f \,\in\, \mathcal{S}^\perp,\ \|f\| \le 1} f(y_0)\,.$$

Corollary 4.2 (Best Affine Approximation)**.** *If $\mathcal{A}$ is an affine subspace and y_0 is an element of a real, normed linear space, then*

$$\inf_{y \,\in\, \mathcal{A}} \|y - y_0\| \;=\; \max_{f \,\in\, (\mathcal{A} - a)^\perp,\ \|f\| \le 1} f(y_0 - a)$$

in which a is any element of $\mathcal{A}$.

Proof. Replace Theorem 4.1's y, y_0, $\mathcal{S}$ by $y - a$, $y_0 - a$, $\mathcal{A} - a$. □

Corollary 4.3. *Let $T : \mathbb{R}^m \to \mathbb{R}^p$ be a linear transformation. For every $y_0 \in \mathbb{R}^m$, each optimization problem below is well-defined if and only if $h \in T(\mathbb{R}^m)$, in which case the optimal values are equal.*

$$\min_{y \,\in\, \mathbb{R}^m \,:\, Ty = h} \|y - y_0\| \;=\; \max_{g \,\in\, (\mathbb{R}^n)^* \,:\, \|T^* g\| \le 1} g(Ty_0 - h)$$

Proof. Only the well-posedness of the maximization needs to be considered. If $h \in T(\mathbb{R}^m)$, then $h = Tu$ for some u, so the objective function,

$$g(Ty_0 - h) = gT(y_0 - u) = (T^*g)(y_0 - u) \le \|T^*g\|\, \|y_0 - u\| \le \|y_0 - u\|\,,$$

is bounded above for every $g \in (\mathbb{R}^n)^*$. The maximum is attained because the feasible set is closed in a finite dimensional space.

Conversely, suppose the maximization is well posed. If $g \in T(\mathbb{R}^n)^\perp = \ker(T^*)$, then g and all its multiples are feasible. Hence $g(h) = 0$, lest by scaling g it would be possible to make $g(Ty_0 - h) = g(h)$ arbitrarily large. Thus $h \in {}^\perp[T(\mathbb{R}^m)^\perp] = T(\mathbb{R}^m)$ by the usual identification of $\mathbb{R}^m$ and $(\mathbb{R}^m)^{**}$.

All that remains is to establish the equality using Corollary 4.2. Choose $\mathcal{A} = \{y \in \mathbb{R}^m : Ty = h\}$ and $a \in \mathcal{A}$. Now $\mathcal{A} - a = \ker(T)$, so

$$(\mathcal{A} - a)^\perp = [\ker(T)]^\perp = [{}^\perp(T^*(\mathbb{R}^n)^*)]^\perp = T^*(\mathbb{R}^n)^* \,.$$

This means $f \in (\mathcal{A} - a)^\perp$ if and only if $f = T^* g$ for some $g \in (\mathbb{R}^n)^*$. Thus Corollary 4.2's maximization is over all such g with $\|T^* g\| = \|f\| \le 1$. Finally, the objective function is

$$f(y_0 - a) = (T^* g)(y_0 - a) = gT(y_0 - a) = g(Ty_0 - Ta) = g(Ty_0 - h)$$

as desired. □

Theorem 4.4 ($(P_\bullet)_{\min} \Leftrightarrow (P_\bullet)_{\max}$).

- *In addition to Hypothesis 2.5, suppose that $F(y_0, x_0) = 0$, and that $D_1 F(y_0, x_0)$ is onto.*

⇒ *There is a neighborhood of x_0 where Table 4's problems $(P_\mathcal{T})_{\min}$ and $(P_\mathcal{T})_{\max}$ are well-defined and their values are equal, and similarly for the (P_0) and (P_ℓ) pairs of dual problems.*

Proof. By Lemma 3.7, $D_1 F(y_0, x)$ is onto for every $x \in K_{x_0}^{(3.7)}$. Therefore by Corollary 4.3 the following problems are well-defined and their optimal values are equal for every $h \in \mathbb{R}^p$.

$$\min_{y \,:\, D_1 F(y_0,x) y \,-\, h \,=\, 0} \|y - y_0\| \;=\; \max_{f \,:\, \|D_1 F(y_0,x)^* f\| \,\le\, 1} f(D_1 F(y_0, x) y_0 - h)$$

Choosing $h = D_1 F(y_0, x) y_0 - F(y_0, x)$ gives the theorem's conclusion for the $(P_\mathcal{T})$ dual problems.

In particular $D_1 F(y_0, x_0)$ is onto, so also by Corollary 4.3 the following problems are well-defined and their optimal values are equal for every $h \in \mathbb{R}^p$.

$$\min_{y \,:\, D_1 F(y_0,x_0) y \,-\, h \,=\, 0} \|y - y_0\| \;=\; \max_{f \,:\, \|D_1 F(y_0,x_0)^* f\| \,\le\, 1} f(D_1 F(y_0, x_0) y_0 - h)$$

The choice $h = D_1 F(y_0, x_0) y_0 - F(y_0, x)$ gives the conclusion for the (P_0) dual problems; similarly $h = D_1 F(y_0, x_0) y_0 - D_2 F(y_0, x_0)(x - x_0)$ for the (P_ℓ) problems. □

5. Second Equivalence, $(P_\mathcal{T})_{\max} \Leftrightarrow (P_0)_{\max}$

The second equivalence to be proved (in the notation of Table 4),

$$\mu_\mathcal{T}(x)_{\max} \;\underset{x_0}{\dot{\simeq}}\; \mu_0(x)_{\min} \,,$$

says that the feasible set $\{f : \|D_1 F(y_0, x)^* f\| \le 1\}$ can be replaced by one that is independent of x. The proof of this is self-contained.

Theorem 5.1 ($(P_\mathcal{T})_{\max} \Leftrightarrow (P_0)_{\max}$).

- *In addition to Hypothesis 2.5, suppose that $F(y_0, x_0) = 0$, and that $D_1 F(y_0, x_0)$ is onto.*

$\Rightarrow$ *There is a neighborhood of x_0 where both of Table 4's optimization problems $(P_{\mathcal{T}})_{\max}$ and $(P_0)_{\max}$ are well-defined. Their values are rationally equivalent at x_0 in the sense of Definition 2.3.*

Proof. The hypotheses suffice to invoke Theorem 4.4 which says $(P_{\mathcal{T}})_{\max}$ and $(P_0)_{\max}$ are well-defined on some neighborhood $N^{(1)}$ of x_0. Let $\mathcal{C}(x)$ be the feasible set in these maximizations.

$$\begin{aligned}
\mu_{\mathcal{T}}(x)_{\max} &= \max_{f \,\in\, \mathcal{C}(x)} f(F(y_0, x)) \\
\mu_0(x)_{\max} &= \max_{f \,\in\, \mathcal{C}(x_0)} f(F(y_0, x)) \\
\text{where } \mathcal{C}(x) &= \{f \in (\mathbb{R}^m)^* : \|D_1 F(y_0, x)^* f\| \le 1\}
\end{aligned}$$

The proof has three steps that culminate in equations (14), (15) and (16), respectively.

(Step 1.) If $f_1 \in \operatorname{bd}(\mathcal{C}(x_0)) = \{f : \|D_1 F(y_0, x_0)^* f\| = 1\}$, then

$$\begin{aligned}
\big|\|D_1 F(y_0, x)^* f_1\| - 1\big| &= \big|\|D_1 F(y_0, x)^* f_1\| - \|D_1 F(y_0, x_0)^* f_1\|\big| \\
&\le \|D_1 F(y_0, x)^* f_1 - D_1 F(y_0, x_0)^* f_1\| \\
&\le \|D_1 F(y_0, x)^* - D_1 F(y_0, x_0)^*\| \, \|f_1\| \qquad (13) \\
&= \|D_1 F(y_0, x) - D_1 F(y_0, x_0)\| \, \|f_1\| \\
&\le \|D_1 F(y_0, x) - D_1 F(y_0, x_0)\| \max_{f \,\in\, \operatorname{bd}(\mathcal{C}(x_0))} \|f\| \,.
\end{aligned}$$

The linear transformation $D_1 F(y_0, x_0)$ is onto, so its adjoint $D_1 F(y_0, x_0)^*$ is one-to-one. Hence $\|D_1 F(y_0, x)^* f\|$ defines a norm on the dual space whose closed unit ball is $\mathcal{C}(x_0)$. Thus, in the last of equation (13)'s bounds, the maximum is finite because $\mathcal{C}(x_0)$ is compact. There also, the difference term converges to 0 as $x \to x_0$ because F is continuously differentiable. Altogether, $\|D_1 F(y_0, x)^* f_1\|$ converges to 1 uniformly on $\operatorname{bd}(\mathcal{C}(x_0))$ as $x \to x_0$. This means, for every $\epsilon > 0$, there is a neighborhood $N^{(2)}(\epsilon)$ of x_0, such that

$$x \in N^{(2)}(\epsilon) \text{ and } f_1 \in \operatorname{bd}(\mathcal{C}(x_0)) \;\Rightarrow\; 1 - \epsilon \le \|D_1 F(y_0, x)^* f_1\| \le 1 + \epsilon \qquad (14)$$

(Step 2.) Choose $x \in N^{(1)} \cap N^{(2)}(\epsilon)$, and then choose any nonzero $f \in \mathcal{C}(x)$, and finally let $f_1 = f / \|D_1 F(y_0, x_0)^* f\| \in \operatorname{bd}(\mathcal{C}(x_0))$. Assume without loss of generality that $\epsilon < 1$. It is now possible to calculate

$$\begin{aligned}
\|(1-\epsilon) D_1 F(y_0, x_0)^* f\| &= (1-\epsilon)\, \|D_1 F(y_0, x_0)^* f\| \\
&\le \|D_1 F(y_0, x)^* f_1\| \, \|D_1 F(y_0, x_0)^* f\| \\
&= \|D_1 F(y_0, x)^* f\| \\
&\le 1\,,
\end{aligned}$$

in which the first inequality is from equation (14), and the second is because $f \in \mathcal{C}(x)$. This proves $(1-\epsilon) f \in \mathcal{C}(x_0)$.

Similarly, choose any nonzero $f_0 \in \mathcal{C}(x_0)$ and let $f_1 = f_0 / \|D_1 F(y_0, x_0)^* f_0\| \in \operatorname{bd}(\mathcal{C}(x_0))$. It now follows that

$$\begin{aligned} \|(1+\epsilon)^{-1} D_1 F(y_0, x)^* f_0\| &= (1+\epsilon)^{-1} \|D_1 F(y_0, x)^* f_0\| \\ &= (1+\epsilon)^{-1} \|D_1 F(y_0, x)^* f_1\| \, \|D_1 F(y_0, x_0)^* f_0\| \\ &\le \|D_1 F(y_0, x_0)^* f_0\| \\ &\le 1 \end{aligned}$$

in which the first inequality is again from equation (14), but now the second is because $f_0 \in \mathcal{C}(x_0)$. This proves $(1+\epsilon)^{-1} f_0 \in \mathcal{C}(x)$.

These two calculations establish the next implication.

$$x \in N^{(1)} \cap N^{(2)}(\epsilon) \;\Rightarrow\; (1-\epsilon)\,\mathcal{C}(x) \;\subseteq\; \mathcal{C}(s_0) \;\subseteq\; (1+\epsilon)\,\mathcal{C}(x) \tag{15}$$

(Step 3.) Choose $x \in N^{(1)} \cap N^{(2)}(\epsilon)$, and then choose $f_{\mathcal{T}} \in \mathcal{C}(x)$ that attains $\mu_{\mathcal{T}}(x)_{\max}$. Equation (15) asserts $(1-\epsilon) f_{\mathcal{T}} \in \mathcal{C}(x_0)$, so

$$\mu_0(x)_{\max} \;=\; \max_{f \,\in\, \mathcal{C}(x_0)} f(F(y_0, x)) \;\ge\; (1-\epsilon) f_{\mathcal{T}}(F(y_0, x)) \;=\; (1-\epsilon)\, \mu_{\mathcal{T}}(x)_{\max}\,.$$

Similarly, choose $f_0 \in \mathcal{C}(x_0)$ that attains $\mu_0(x)_{\max}$. Now equation (15) asserts $(1+\epsilon)^{-1} f_0 \in \mathcal{C}(x)$, so

$$\mu_{\mathcal{T}}(x)_{\max} \;=\; \max_{f \,\in\, \mathcal{C}(x)} f(F(y_0, x)) \;\ge\; (1+\epsilon)^{-1} f_0(F(y_0, x)) \;=\; (1+\epsilon)^{-1} \mu_0(x)_{\max}\,.$$

Together these two inequalities provide the final implication,

$$x \in N^{(1)} \cap N^{(2)}(\epsilon) \;\Rightarrow\; (1-\epsilon)\, \mu_{\mathcal{T}}(x)_{\max} \le \mu_0(x)_{\max} \le (1+\epsilon) \mu_{\mathcal{T}}(x)_{\max}\,, \tag{16}$$

which is Definition 2.3's equation (4). □

6. Applications

6.1. **Canonical Estimates for Backward Errors.** For the implicitly defined numerical problem $F(y, x) = 0$ with data y and solution x, Section 1.2 formulates the minimal size of the backward error as

$$\mu(x) = \min_{y \,:\, F(y, x) \,=\, 0} \|y - y_0\|\,,$$

where y_0 is some data for which x approximates a solution x_0, whose precise value is unknown. Since $\mu(x)$ can be evaluated by solving this minimization problem, which is independent of x_0, the backward error $y - y_0$ can be had without knowing the actual error, $x - x_0$. Thus in principal the accuracy of numerical calculations can be assessed in von Neumann's sense based on the expectation of having available the minimal size of the backward error.

There are impediments to realizing this approach. Expressions for $\mu(x)$ are not easily obtained, in part because the constraints $F(y, x) = 0$ may be nonlinear, and even if obtained they may be difficult to evaluate, or interpret. For these reasons several of the papers cited in Table 3 instead derive estimates for $\mu(x)$ using reasoning specific to the numerical problems they consider.

This section suggests two estimates for $\mu(x)$ that apply to all numerical problems. First, Table 4 indicates that $\mu(x) \simeq^{\partial}_{x_0} \mu_{\mathcal{T}}(x)$. This value is guaranteed to accurately estimate the minimal size of the backward error in both the differential and the rational senses of Definition 2.1 and 2.3. Moreover, many algorithms are available to solve the linearly constrained best approximation problem for $\mu_{\mathcal{T}}(x)$.

Theorem 6.1 (Computable Backward Error Estimate)**.**

- *In addition to Hypothesis 2.5, suppose that $F(y_0, x_0) = 0$, and that $D_1 F(y_0, x_0)$ is onto.*

$\Rightarrow$ *For the numerical problem $F(y, x) = 0$ with data y_0 and exact solution x_0, there is a neighborhood of x_0 in which the size of the minimal backward error corresponding to an approximate solution $x \approx x_0$ can be estimated by solving the linearly constrained optimization problems*

$$\begin{aligned} \mu_{\mathcal{T}}(x) &= \min_{\Delta y \,:\, D_1 F(y_0,x)\Delta y \,=\, F(y_0,x)} \|\Delta y\| \\ &= \max_{f \,:\, \|D_1 F(y_0,x)^* f\| \,\le\, 1} f(F(y_0,x))\,. \end{aligned}$$

This value estimates the minimal size of the backward error in both the differential and rational senses of Definitions 2.1 and 2.3.

Proof. This combines Theorems 3.11, 4.4 and Corollary 2.8. □

The second estimate explains what about a numerical problem governs the minimal size of its backward errors. The explanation involves a class of norms that has been used in already the proof of Theorem 5.1. If a linear transformation $T : \mathbb{R}^m \to \mathbb{R}^p$ is onto, then its adjoint T^* is one-to-one, so $\|T^* f\|$ defines a norm on the dual space, $(\mathbb{R}^p)^*$. The dual of this norm, viewed as a norm on $\mathbb{R}^p$, is given by the following construction.

Lemma 6.2. *If a linear transformation $T : \mathbb{R}^m \to \mathbb{R}^p$ is onto, then*

$$\|v\|_T := \max_{f \,:\, \|T^* f\| \,\le\, 1} f(v)\,,$$

is a norm on $\mathbb{R}^p$

All of Table 4's estimates for $\mu(x)$ can be expressed in terms of Lemma 6.2's notation,

$$\begin{aligned} \mu_{\mathcal{T}}(x) &= \|F(y_0,x)\|_{D_1 F(y_0,x)} \\ \mu_0(x) &= \|F(y_0,x)\|_{D_1 F(y_0,x_0)} \\ \mu_\ell(x) &= \|D_2 F(y_0,x_0)(x - x_0)\|_{D_1 F(y_0,x_0)}\,. \end{aligned}$$

The expression for $\mu_0(x)$ is special because its norm, $\|\cdot\|_{D_1 F(y_0,x_0)}$, is independent of x, and because the quantity inside, $F(y_0, x)$, is the numerical problem's residual. Thus, up to a first-order differential approximation, the minimal size of the backward error is simply a norm of the residual. It often happens that this norm is unique.

Theorem 6.3 (Canonical Backward Error Estimate)**.**

- *In addition to Hypothesis 2.5, suppose that $F(y_0, x_0) = 0$, and suppose that both $D_1 F(y_0, x_0)$ and $D_2 F(y_0, x_0)$ are onto.*

$\Rightarrow$ *The only member of $\mu(x)$'s rational equivalence class that is an x-invariant norm of $F(y_0, x)$ is $\mu_0(x)$.*

Proof. Suppose $\mu(x) \simeq^{\div}_{x_0} \|F(y_0, x)\|_0$ for some norm, $\|\cdot\|_0$, that does not depend on x. Since $\simeq^{\div}_{x_0}$ is an equivalence relation, and $\mu_0(x) \simeq^{\div}_{x_0} \mu(x)$ by Theorems 3.11

and 6.4, therefore $\|F(y_0,x)\|_{D_1 F(y_0,x_0)} = \mu_0(x) \simeq^{\div}_{x_0} \|F(y_0,x)\|_0$. This means, for every $\epsilon > 0$ there is a neighborhood $N_{x_0}(\epsilon)$ of x_0, such that for every $x \in N_{x_0}(\epsilon)$,

$$(1-\epsilon)\, \|F(y_0,x)\|_0 \;\le\; \|F(y_0,x)\|_{D_1 F(y_0,x_0)} \;\le\; (1+\epsilon)\, \|F(y_0,x)\|_0 \,. \tag{17}$$

By hypothesis $D_2 F(y_0,x_0) : \mathbb{R}^n \to \mathbb{R}^p$ is onto, so there is a subspace of $\mathbb{R}^n$ to which the restriction of this linear transformation is an isomorphism. Moreover, $\mathbb{R}^n$ can be represented as the sum of this subspace and any complementary subspace. Without loss of generality then, assume that F is a function with domain $\mathcal{D} \subseteq \mathbb{R}^m \times \mathbb{R}^{n-p} \times \mathbb{R}^p$ for which $F(y_0,w_0,v_0) = 0$ and $D_3 F(y_0,w_0,v_0)$ is an isomorphism. In this notation, the former variables "x" are now "(w,v)."

The inversion theorem says the function defined by $f(v) = F(y_0,w_0,v)$ is an homeomorphism between a neighborhood of v_0 and a neighborhood, N_0, of $F(y_0,w_0,v_0) = 0$. If $u \in \mathbb{R}^p$, then $\alpha u \in N_0$ for all sufficiently small α, because N_0 is a neighborhood of 0. Also $(w_0, f^{-1}(\alpha u)) \in N_{x_0}(\epsilon)$ for all sufficiently small α, because $N_{x_0}(\epsilon)$ is a neighborhood of $x_0 = (w_0,v_0) = (w_0, f^{-1}(0))$.

Choose $\alpha \neq 0$ for which both $\alpha u \in N_0$ and $(w_0, f^{-1}(\alpha u)) \in N_{x_0}(\epsilon)$. The latter means that equation (17) can be applied. Since $F(y_0,w_0,f^{-1}(\alpha u)) = f(f^{-1}(\alpha u)) = \alpha u$, therefore equation (17) is

$$(1-\epsilon)\, \|\alpha u\|_0 \;\le\; \|\alpha u\|_{D_1 F(y_0,x_0)} \;\le\; (1+\epsilon)\, \|\alpha u\|_0 \,.$$

Hence $\|u\|_0 = \|u\|_{D_1 F(y_0,x_0)}$ because ϵ is arbitrary and $\alpha \neq 0$. Hence $\|\cdot\|_0 = \|\cdot\|_{D_1 F(y_0,x_0)}$ because u is arbitrary. □

6.2. **Third Equivalence and Directional Derivatives.**

For the distance between a point $y_0 \in \mathbb{R}^n$ and a parameterized set $S(x) = \{y : F(y,x) = 0\}$, the directional differentiability with respect to the parameter x follows from Table 4's problem $(P_\ell)_{\max}$. Before discussing differentiability it is therefore necessary to establish the equivalence of this problem to the table's others, thereby completing the table.

Theorem 6.4 $((P_0)_{\max} \Leftrightarrow (P_\ell)_{\max})$**.**

- *In addition to Hypothesis 2.5, suppose that* $F(y_0,x_0) = 0$, *and that* $D_1 F(y_0,x_0)$ *is onto.*

⇒ *There is a neighborhood of* x_0 *where both of Table 4's optimization problems* $(P_0)_{\max}$ *and* $(P_\ell)_{\max}$ *are well-defined. Their values are differentially equivalent at* x_0 *in the sense of Definition 2.1.*

Proof. Let $\|\cdot\|_T$ be Lemma 6.2's norm for the linear transformation $T = D_1 F(y_0,x_0)$. Let $\mathcal{T}_{x_0}(y,x) = D_2 F(y,x_0)(x - x_0) + F(y,x_0)$ be the linear function parameterized by y whose graph is tangent to $F(y,x)$'s at $x = x_0$. (Note this is not the $\mathcal{T}_{y_0}$ of Definition 2.5.) In this notation, $\mu_0(x)_{\max} = \|F(y_0,x)\|_T$ and $\mu_\ell(x)_{\max} = \|\mathcal{T}_{x_0}(y_0,x)\|_T$. Thus by the triangle inequality,

$$\left|\mu_0(x)_{\max} - \mu_\ell(x)_{\max}\right| \le \|F(y_0,x) - \mathcal{T}_{x_0}(y_0,x)\|_T \,,$$

so

$$\lim_{x \to x_0} \frac{|\mu_0(x)_{\max} - \mu_\ell(x)_{\max}|}{\|x - x_0\|} \;\le\; \lim_{x \to x_0} \frac{\|F(y_0,x) - \mathcal{T}_{x_0}(y_0,x)\|_T}{\|x - x_0\|} \;=\; 0\,,$$

because F is continuously Fréchet differentiable. □

The directional differentiability, with respect to x, of the distance between y_0 and $S(x)$ is established in the following sense.

Definition 6.5 (Fréchet Directional Derivative [47, p. 479, eqn. 3])**.** Let $f(x)$ be a function among normed linear spaces, and let $h(x)$ be a positive homogeneous function among the same. That is, $h(tx) = th(x)$ when $t \geq 0$. If

$$\lim_{\Delta x \to 0} \frac{\| f(x_0 + \Delta x) - f(x_0) - h(\Delta x) \|}{\| \Delta x \|} = 0 \,,$$

then f is directionally differentiable at x_0 with derivative $h(\Delta x)$ in direction Δx. Shapiro [47] discusses variations of this and other definitions (Gâteaux, Hadamard).

The following theorem extends equation (1) to arbitrary norms under weaker differential hypotheses than those of [6, pp. 434, theorem 5.42].

Theorem 6.6 (Directional Differentiability)**.** *Assume*

- $\mathcal{D} \subseteq \mathbb{R}^m \times \mathbb{R}^n$ *is a neighborhood of* (y_0, x_0)*,*
- $F : \mathcal{D} \to \mathbb{R}^p$ *is continuously differentiable,*
- $F(y_0, x_0) = 0$*,*
- $D_1 F(y_0, x_0) \in \hom(\mathbb{R}^m, \mathbb{R}^p)$ *is onto.*

Let $\mu(x)$ *be the distance from* y_0 *to the set* $S(x) = \{y : F(y, x) = 0\}$*,*

$$\mu(x) = \min_{y \,:\, F(y,x) \,=\, 0} \| y - y_0 \| \,,$$

and let $\Delta x \in \mathbb{R}^n$ *be a vector. The function* $\mu(x)$ *is Fréchet directionally differentiable at* x_0*. The derivative in the direction* Δx *is*

$$\mu_\ell(x_0 + \Delta x) = \min_{\Delta y \,:\, DF(y_0, x_0)(\Delta y, \Delta x) \,=\, 0} \| \Delta y \| \,. \tag{18}$$

Proof. Let $\|\cdot\|_T$ be Lemma 6.2's norm for the linear transformation $T = D_1 F(y_0, x_0)$. With this notation and by Theorem 4.4, equation (18)'s function $\mu_\ell(x_0 + \Delta x)$ can be expressed as $\| D_2 F(y_0, x_0) \Delta x \|_T$. This is a "positive homogeneous function" of Δx in Definition 6.5's terminology. Moreover, $\mu(x) \simeq^{\partial}_{x_0} \mu_\ell(x)$ by all the theorems (principally 3.11, 5.1, and 6.4) that establish Table 4's equivalences. For $x = x_0 + \Delta x$, since $\mu(x_0) = 0$, the definition of the differential equivalence $\mu(x) \simeq^{\partial}_{x_0} \mu_\ell(x)$ in Definition 2.1's equation (3) is exactly Definition 6.5's limit. □

7. Conclusion

The first-order sensitivity of a metric projection has been studied, for a set defined by parameterized equality constraints, by altering the set in ways that keep the optimum-value function in an equivalence class of functions that have identical first-order differential properties. The utility of this approach has been demonstrated by estimating optimal backward errors in numerical analysis, and by evaluating the directional derivatives of the distance between a point and a perturbable set under weaker differential hypotheses than previously considered in optimization theory.

Whether the method of performing sensitivity analyses in terms of equivalence relations can be applied more generally suggests the following questions. The first regards a topic not treated in this paper, the parameterized solution function.

1. For the problems in Table 4, let $y_*(x)$ attain the minimum $\mu_*(x)$. Is it possible to establish differential equivalences among the solution functions, $y_*(x)$? This is not trivial because Table 1 shows that $\operatorname{dist}(y_0, S)$ may be differentiable (in the table's case, with respect to y_0) when $P_S(y_0)$ is not.

The next questions regard further weakening the hypotheses on the function F. This addresses the generality of the proofs used in this paper.

(2) The partial derivatives of F with respect to the parametric variable x, $D_2 F$, appear only in the final row of Table 4. Can equivalences among the functions in the other rows be obtained if F is just continuous with respect to x?

(3) Can Table 4's blanket assumption that $D_1 F(y_0, x_0)$ is onto be relaxed for any of the equivalences?

The greatest interest lies in establishing equivalences among the values of optimization problems other than metric projections, onto sets defined with other than equality constraints, and at points x_0 other than where the optimal value vanishes.

(4) Can equivalences be established among the minima in Table 4 if their common objective function, $\|y - y_0\|$, is replaced by:
  (a) $f(y) - f(y_0)$, or perhaps $\|f(y) - f(y_0)\|$, for some function f,
  (b) $f(y, x) - f(y_0, x)$, or perhaps $\|f(y, x) - f(y_0, x)\|$,

(5) What optimization problems are equivalent to

$$\min_{y \,:\, F(y,x) \in C} \|y - y_0\| ,$$

where C is closed and convex? When this construction enforces simple inequality constraints then it reduces to the equality constraints treated in this paper provided only active constraints need be considered.

(6) Can equivalences at x_0 be established among the minima in Table 4 without the hypothesis $F(y_0, x_0) = 0$? In view of the second column in Table 2 it would seem that additional hypotheses may be necessary.

Equivalence relations among functions that enforce various kinds of approximations are the subject of the next four questions.

(7) What exactly is the relationship between the directional and rational equivalences of Definitions 2.1 and 2.3? Note that Corollary 2.8 can be interpreted as asserting that a rational equivalence class for a certain kind of function is contained in a differential equivalence class.

(8) Is there an equivalence relation for functions into a normed linear space analogous to Definition 2.3's rational equivalence for real-valued functions?

(9) Can equivalence relations be used to investigate other perturbational properties? Equivalences for studying second order sensitivities can be found in the papers of Shapiro [44] [46].

(10) Can Theorem 6.4 be improved to assert rational equivalence between the optimal value in the final row of Table 4 and the others?

It would be useful to know if this type of perturbation analysis can be used as well in function spaces.

(11) Can the results of this paper be established in function spaces, that is, in infinite dimensional spaces? In view of Section 1.1 and Table 2, it may be best to begin with Hilbert spaces.

Finally, the parameterized versions of familiar theorems in real analysis that support this approach to perturbation analysis, and in particular the collocation theorem, beg the following questions.

(12) Do the constructions $(y, x) \mapsto (y_T, x)$ or $(y, x) \mapsto (y_f, x)$ in the proof of Lemma 3.10 lead to immediate proofs of parameterized implicit or inverse function theorems?

(13) If the answer to the previous question is affirmative, then what is the nature of the functional relationship between the parameter and the associated implicit function?

References

1. S. Banach, *Théorie des opérations linéaires*, Subwencji Funduszu Kultury Narodowej, Warsaw, 1932, Translated to English in [2].
2. ———, *Theory of linear operators*, North-Holland, Amsterdam, 1987, Translated by F. Jellet from a 1979 reprinting of [1].
3. S. G. Bartels and D. J. Higham, *The structured sensitivity of Vandermonde-like systems*, Numerische Mathematik **62** (1992), no. 1, 17–33.
4. R. G. Bartle, *The elements of real analysis*, second ed., John Wiley & Sons, New York, 1976.
5. J. F. Bonnans and A. Shapiro, *Optimization problems with perturbations: a guided tour*, SIAM Review **40** (1998), no. 2, 228–264.
6. ———, *Perturbation analysis of optimization problems*, Springer Series in Optimization Research, Springer-Verlag, New York, 2000.
7. J. R. Bunch, J. W. Demmel, and C. F. Van Loan, *The strong stability of algorithms for solving symmetric linear systems*, SIAM Journal on Matrix Analysis and Applications **10** (1989), no. 4, 494–499.
8. S. Chandrasekaran and I. C. F. Ipsen, *Backward errors for eigenvalue and singular value decompositions*, Numerische Mathematik **68** (1994), no. 2, 215–223.
9. F. H. Clarke, R. J. Stern, and P. R. Wolenski, *Proximal smoothness and the lower-C^2 property*, Journal of Convex Analysis **2** (1995), 117–144.
10. A. J. Cox and N. J. Higham, *Backward error bounds for constrained least squares problems*, BIT **39** (1999), no. 2, 210–227.
11. A. V. Fiacco and A. Ghaemi, *Preliminary sensitivity analysis of a stream pollution abatement system*, Mathematical Programming with Data Perturbations I (A. V. Fiacco, ed.), Lecture Notes in Pure and Applied Mathematics, vol. 73, Marcel Dekker, Inc., New York, 1982, Papers presented at the First Symposium on Mathematical Programming with Data Perturbations, George Washington University, May 24–25, 1979., pp. 111–130.
12. S. Fitzpatrick and R. R. Phelps, *Differentiability of the metric projection in Hilbert space*, Transactions of the American Mathematical Society **270** (1982), no. 2, whole number 563, 483–501.
13. V. Frayssé and V. Toumazou, *A note on the normwise perturbation theory for the regular generalized eigenproblem $Ax = \lambda Bx$*, Numerical Linear Algebra With Applications **5** (1998), 1–10.
14. H. H. Goldstine, *The computer from Pascal to von Neumann*, Princeton University Press, Princeton, 1972.
15. L. M. Graves, *Some mapping theorems*, Duke Mathematical Journal **17** (1950), 111–114.
16. J. F. Grcar, *A matrix lower bound*, Technical Report LBNL-50635, Lawrence Berkeley National Laboratory, 2002, Submitted for publication.
17. M. Gu, *Backward perturbation bounds for linear least squares problems*, SIAM Journal on Matrix Analysis and Applications **20** (1998), no. 2, 363–372.
18. A. Haraux, *How to differentiate the projection on a convex set in Hilbert space. Some applications to variational inequalities*, Journal of the Mathematical Society of Japan **29** (1975), no. 4, 615–632.
19. D. J. Higham and N. J. Higham, *Backward error and condition of structured linear systems*, SIAM Journal on Matrix Analysis and Applications **13** (1992), no. 1, 162–175.
20. ———, *Componentwise perturbation theory for linear systems with multiple right-hand sides*, Linear Algebra and Its Applications **174** (1992), 111–129.
21. ———, *Structured backward error and condition of generalized eigenvalue problems*, SIAM Journal on Matrix Analysis and Applications **20** (1998), no. 2, 493–512.

22. N. J. Higham, *Accuracy and stability of numerical algorithms*, Society for Industrial and Applied Mathematics, Philadelphia, 1996.
23. J.-B. Hiriart-Urruty, *At what points is the projection mapping differentiable?*, American Mathematical Monthly **89** (1982), no. 7, 456–460.
24. R. B. Holmes, *Smoothness of certain metric projections on Hilbert space*, Transactions of the American Mathematical Society **183** (1973), no. 457, 87–100.
25. R. Karlson and B. Waldén, *Estimation of optimal backward perturbation bounds for the linear least squares problem*, BIT **37** (1997), no. 4, 862–869.
26. J. B. Kruskal, *Two convex counterexample: a discontinuous envelope function and a nondifferentiable nearest-point mapping*, Proceedings of the American Mathematical Society **23** (1969), no. 3, 697–703.
27. E. S. Levitin, *Perturbation theory in mathematical programming and its applications*, Wiley-Interscience Series in Discrete Mathematics and Optimization, John Wiley & Sons, Chichester, 1994.
28. D. G. Luenberger, *Optimization by vector space methods*, John Wiley & Sons, New York, 1969.
29. A. N. Malyshev, *Optimal backward perturbation bounds for the LSS problem*, BIT **41** (2001), no. 3, 430–432.
30. A. N. Malyshev and M. Sadkane, *Computation of optimal backward perturbation bounds for large sparse linear least squares problems*, BIT **41** (2002), no. 4, 739–747.
31. G. P. McCormick and R. Tapia, *The gradient projection method under mild differentiability conditions*, SIAM J. Control **10** (1972), 93–98.
32. F. Mignot, *Contrôle dans les inéquations variationelles elliptiques*, Journal of Functional Analysis **22** (1976), no. 2, 130–185.
33. J. Moreau, *Proximité et dualité dans un espace hilbertian*, Bulletin de la Société mathématique de France **93** (1965), no. 3, 273–299.
34. J. von Neumann and H. H. Goldstine, *Numerical inverting of matrices of high order*, Bulletin of the American Mathematical Society **53** (1947), no. 11, 1021–1099, Reprinted in [60, v. 5, pp. 479–557].
35. W. Oettli and W. Prager, *Compatibility of approximate solution of linear equations with given error bounds for coefficients and right-hand sides*, Numerische Mathematik **6** (1964), 405–409.
36. J. M. Ortega and W. C. Rheinboldt, *Iterative solution of nonlinear equations in several variables*, Academic Press, New York, 1970.
37. R. R. Phelps, *Convex sets and nearest points, I*, Proceedings of the American Mathematical Society **8** (1957), 790–797.
38. ______, *Convex sets and nearest points, II*, Proceedings of the American Mathematical Society **8** (1958), 867–873.
39. ______, *Metric projections and the gradient projection method in Banach spaces*, SIAM Journal on Control and Optimization **23** (1985), no. 6, 973–977.
40. R. A. Poliquin, R. T. Rockafellar, and L. Thibault, *Local differentiability of distance functions*, Transactions of the American Mathematical Society **352** (2000), no. 11, whole number 786, 5231–5249.
41. J. L. Rigal and J. Gaches, *On on the compatibility of a given solution with the data of a linear system*, Journal of the Association of Computing Machinery **14** (1967), no. 3, 543–548.
42. S. M. Robinson, *Stability theory for systems of inequalities. Part II: Differentiable nonlinear systems*, SIAM Journal on Numerical Analysis **13** (1976), no. 4, 497–513.
43. A. Shapiro, *Second-order derivatives of extremal-value functions and optimality conditions for semi-inifinite programs*, Mathematics of Operations Research **10** (1985), no. 2, 207–219.
44. ______, *Second order sensitivity analysis and asymptotic theory of parameterized nonlinear programs*, Mathematical Programming **33** (1985), no. 3, 280–299.
45. ______, *On differentiability of metric projections in $\mathbb{R}^n$, 1: Boundary case*, Proceedings of the American Mathematical Society **99** (1987), no. 1, 123–128.
46. ______, *Sensitivity analysis of nonlinear programs and differentiability properties of metric projections*, SIAM Journal on Control and Optimization **26** (1988), no. 3, 628–645.
47. ______, *On concepts of directional differentiability*, Journal of Optimization Theory and Applications **66** (1990), no. 3, 477–487.

48. ———, *Directionally nondifferentiable metric projections*, Journal of Optimization Theory and Applications **81** (1994), no. 1, 203–204.
49. ———, *Existence and differentiability of metric projections in Hilbert space*, SIAM Journal on Optimization **4** (1994), no. 1, 130–141.
50. A. Smoktunowicz, *A note on the strong componentwise stability of algorithms for solving symmetric linear systems*, Demonstratio Mathematica **28** (1995), no. 2.
51. ———, *The strong stability of algorithms for solving the symmetric eigenproblem*, Manuscript, 1995, Cited in [21].
52. G. W. Stewart, *Backward error bounds for approximate Krylov subspaces*, Numerical Linear Algebra With Applications **340** (2002), 81–86.
53. J.-G. Sun, *Perturbation bounds for the Choleski and QR factorizations*, BIT **31** (1991), no. 2, 341–352.
54. ———, *Backward perturbation analysis of certain characteristic subspaces*, Numerische Mathematik **65** (1993), no. 3, 357–382.
55. ———, *A note on backward perturbations for the Hermitian eigenvalue problem*, BIT **35** (1995), no. 4, 385–393.
56. ———, *Optimal backward perturbation bounds for the linear least-squares problem with multiple right-hand sides*, IMA Journal of Numerical Analysis **16** (1996), no. 1, 1–11.
57. ———, *On optimal backward perturbation bounds for the linear least-squares problem*, BIT **37** (1997), no. 1, 179–188.
58. ———, *Bounds for the structured backward errors of Vandermonde systems*, SIAM Journal on Matrix Analysis and Applications **20** (1998), no. 1, 45–59.
59. J.-G. Sun and Z. Sun, *Optimal backward perturbation bounds for underdetermined systems*, SIAM Journal on Matrix Analysis and Applications **18** (1997), no. 2, 393–402.
60. A. H. Taub (ed.), *John von Neumann collected works*, Macmillan, New York, 1963.
61. A. M. Turing, *Rounding-off errors in matrix processes*, The Quarterly Journal of Mechanics and Applied Mathematics **1** (1948), no. 3, 287–308.
62. J. M. Varah, *Backward error estimates for Toeplitz systems*, SIAM Journal on Matrix Analysis and Applications **15** (1994), no. 2, 408–417.
63. B. Waldén, R. Karlson, and J.-G. Sun, *Optimal backward perturbation bounds for the linear least squares problem*, Numerical Linear Algebra With Applications **2** (1995), no. 3, 271–286.
64. J. H. Wilkinson, *Rounding errors in algebraic processes*, Prentice Hall, Englewood Cliffs, New Jersey, 1963.
65. E. H. Zarantonello, *Projections on convex sets in Hilbert space and spectral theory*, Contributions to Nonlinear Functional Analysis (E. H. Zarantonello, ed.), Academic Press, New York, 1971, pp. 237–424.

Lawrence Berkeley National Laboratory, Mail Stop 50A-1148, One Cyclotron Road, Berkeley, CA 94720-8142 USA

E-mail address: jfgrcar@lbl.gov